Data Fusion in the Environmental Domain

Hylke van Der Schaaf, Mike Kobernus, Markus Falgenhauer, Jasmin Pielorz, Kym Watson

Data fusion in the environmental domain

Hylke van der Schaaf[1], Mike Kobernus[2], Markus Falgenhauer[3], Jasmin Pielorz[4], and Kym Watson[1]

[1] Fraunhofer IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany
hylke.vanderschaaf@iosb.fraunhofer.de
kym.watson@iosb.frauhofer.de
[2] NILU - Norsk institutt for luftforskning, Instituttveien 18, Kjeller, Norway
mike.kobernus@nilu.no
[3] Austrian Institute of Technology, Donau-City-Str. 1, 1220 Vienna, Austria
markus.falgenhauer@ait.ac.at
[4] Ubimet GmbH, Dresdner Str. 82, 1200 Vienna, Austria
jpielorz@ubimet.com

Abstract. Air quality and air pollution have a very large impact on human health. The sensitivity to different pollutants varies per person, therefore it is important that citizens can get personalised air quality information. The Personal Environmental Information System (PEIS) aims at delivering just that. The PEIS takes sensor data from several data providers and employs a service-oriented architecture to deliver these observations to the user through a smartphone application. The PEIS also uses scientific models to fuse the sensor data and create new, derived observations. To make scientific models that fuse sensor data fit better in a service-oriented architecture, a software framework called Fusion4Decision was developed. This framework is based on Open Geospatial Consortium standards and allows scientific models written in languages like MATLAB or R to be available as a web service.

Keywords: air quality, open geospatial standards, environmental modelling, fusion

1 Introduction

From the moment of our birth until our very last breath, we spend our lives immersed in a mixture of gases and aerosols known as the air or atmosphere. Unsurprisingly, our well-being is closely connected to the characteristics of the air: some of the gases and chemicals suspended therein are toxic, others may cause allergic reactions, and our bodies are also sensitive to changes in for instance humidity or oxygen levels, air temperature and pressure. Moreover, sensitivity to these substances and changes is very different from one person to another. Despite this, the information on the state of atmosphere available today is entirely generic, and by no means personalized. We intend to address this lack of individualized information through the development of a Personal Environmental Information System (PEIS).

The PEIS aims at enhancing the human to environment interaction by providing personalized air quality, pollen and meteorological data for chosen locations and personalized alerts and notifications about changes in atmospheric conditions [15, 14]. The PEIS also enables users to keep a personal health diary and submit their own environmental observations to the system. Pilot projects currently run in Oslo (Norway) and Vienna (Austria).

The PEIS is a distributed system with a mobile application interface. The data used by the PEIS is supplied by several data providers. Sensor data is available for air quality, meteorology and pollen. These data are presented to the user as they are and they are fed into models. To allow the PEIS to use custom models written in different languages for data fusion, the Fusion4Decision (F4D) system was developed. This allows domain experts to write models in their favourite language and then make those models available in a service oriented architecture, using the OGC SPS standard interface.

This paper provides an example on how heterogeneous data sources and models in the environmental domain can be combined in one language-independent service that agrees with OGC standards [7]. In section 2, we first give an overview over the PEIS mobile application and explain how it relates to the EU project ENVIROFI [2], which is followed by a detailed description of its architecture in section 3. After discussing the different data sources that have to be taken into account in the scope of environmental applications, section 4 lays out how the data fusion service works in case of the PEIS. Finally, we conclude with a discussion of the resulting service and provide a future outlook.

## 2   Personal Environmental Information System

The PEIS is one of three end user applications to come out of ENVIROFI, a co-funded research project within the Future Internet Public Private Partnership (FI-PPP) programme of the EU's Seventh Framework Programme (FP7). The project was dedicated to the environmental usage area of the Future Internet [10]. It explored environmental enablers (applications for collecting and processing environmental data) and provided environmental sector requirements to FIWARE, the FI-PPP core platform project [12, 16]. These were deployed in collaboration with large stakeholder communities in three scenarios, specifically in the terrestrial, atmospheric and marine environments, and laid the foundation for an environmental observation web. In this way, ENVIROFI helps Europe tackle the grand societal challenges of climate change, environmental degradation, and sustainable growth.

The PEIS is an innovative application for smartphones which accesses data and services from multiple providers and combines them into a single system with an intuitive and easy to use interface. It gives users more control over their environment as to what information they track (e.g., particulate matter, ozone, humidity, etc.), how it is presented (GIS overlays as in Fig. 1, tabular data, etc.) and how they are alerted to changes in the environment. The PEIS system provides users with a fully customisable, interactive application that enables the

user to track air borne pollutants, pollen, and meteorology in an entirely personalized manner [15, 14]. Through the creation of profiles, the user can specify just what air quality, pollen and meteorological data should be tracked. Threshold levels can be set for any of these tracked components and notifications triggered when threshold levels are surpassed, alerting the user to the change.
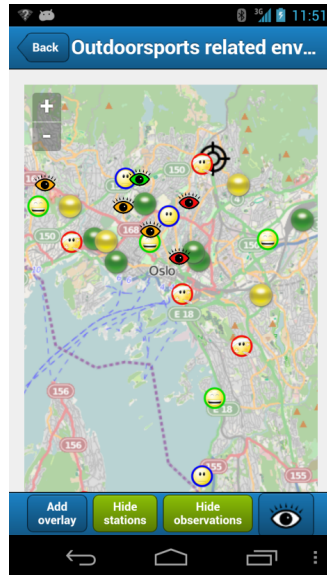
**Fig. 1.** PEIS map-view, showing stations and user observations
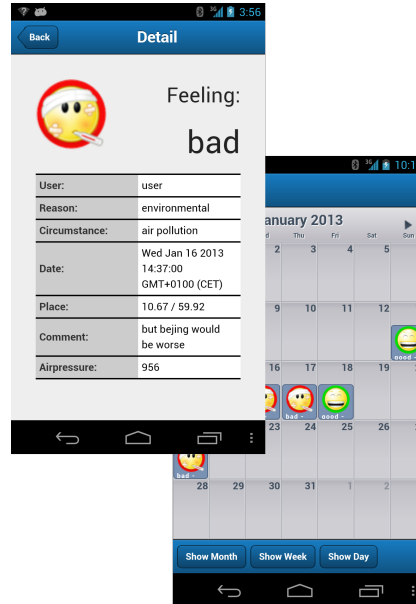
Fig. 2. Health observations and calender view in the PEIS

For most people, this type of system may be more than is required for general, day to day interest in air quality. However, for some, this type of configurable tracking and notification mechanism is key to their health, since sensitivity to airborne pollutants is very different from one person to another and in many cases is highly dependent on the individuals patterns of behaviour or their health. For those who are more at risk the ability to monitor current air quality in a very localized environment is extremely useful. In addition, with air quality forecasts (which use a sophisticated modelling system predicting trends in air quality in the very near future), at risk users can make plans for two or even three days in advance (the system provides up to three days forecasts, with daily updates). This is a far cry from typical air quality data provided to most users which is usually only the current air quality index for the nearest city, often based on the highest relative concentration value measured for any component in the city [5, 13]. In contrast, the PEIS provides air quality data, specific to the users' requirements, in the area where he is currently located, in near real time. Since

the smartphone supplies the user's GPS coordinates, it is possible to not only provide current data for that exact location, but to also store the user's location and create an exposure value for the user every 24 hours. In addition, users can report their own observations, including reports on their current health situation. These are stored in the system, and any additional sensor data available from the phone (such as air pressure, for example) can be stored at the same time. This essentially turns the user into a sensor node in the network. Once the user has been providing their own observations for a while, it is possible to view these in a calendar (Fig. 2), which can aid in spotting potential patterns in the user's health. For example, headaches occurring at times of changing barometric pressure.

3 Architecture

The system is implemented as a service-oriented architecture based on open standards. Central to the architecture is the Mobile Data Acquisition Framework (MDAF [11]), created by the Austrian Institute of Technology (AIT), which manages observations and keeps track of observations relevant to the specific user's needs (Fig. 3).
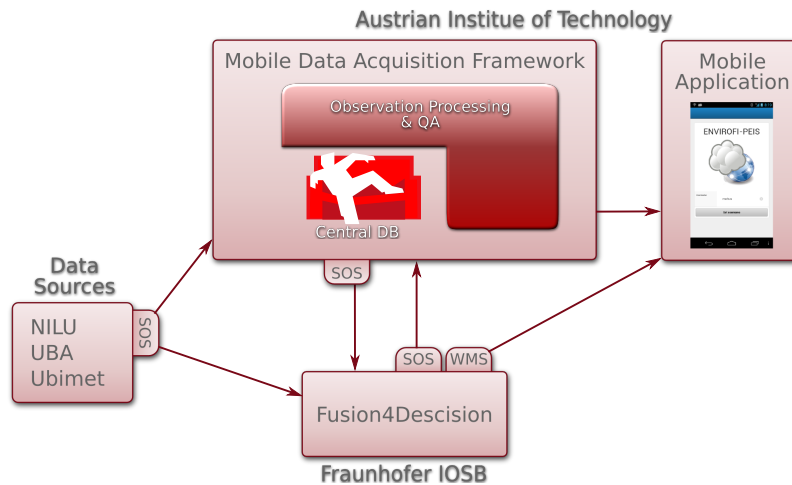
Fig. 3. PEIS Architecture overview
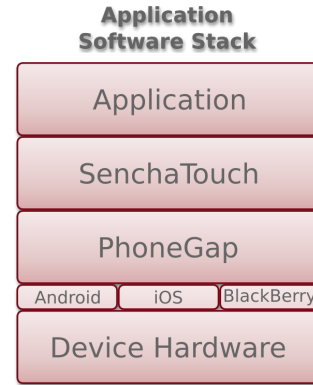
Air quality data, both current and predictive, are provided by the Norwegian Institute for Air Research (NILU) and The Federal Environment Agency (UBA), while meteorological data is provided by Ubimet. Each data provider has its own web services for accessing its data. This data is accessed directly by MDAF, but is also fed into models to create GIS map overlays. These overlays are in turn

accessed by the PEIS using a Web Map Service (WMS). The execution of models is described in more detail in the next section.

Internally MDAF uses GeoCouchDB. CouchDB [1, 6] is a document-based, open source, NoSQL database that combines advantages of an easy to use document-orientated database with the scalability of a relational database. It uses JavaScript as query language and MapReduce with HTTP REST as API. GeoCouchDB is a special development of CouchDB for location aware applications. As documents within the GeoCouchDB we use GeoJSON strings which are JSON strings with additional geographic information.

To enable cross platform compatibility the phone application is developed using Phone-Gap [3] (Fig. 4). Phonegap is a free and open source framework that allows the creation of mobile apps using HTML, CSS and JavaScript and it provides a link to device-specific APIs. The supported platforms are: iPhone, Android, Windows Phone, BlackBerry and many more.

Fig. 4. The PEIS application software stack

On top of Phonegap, we use Sencha Touch [4] which is a flexible HTML5 mobile application framework to enable a quality user experience. It uses the model-view-controller architecture pattern to structure the application.

4 Standardised On-Demand model execution

Part of the core functionality of the PEIS is the processing of data from environmental sensors. These can be both, quantitative as well as qualitative sources, and can be broken down into three basic groups: (i) standardized, gauged sensors, e.g. from professional-grade weather stations measuring meteorological conditions or advanced measurement stations for pollutant concentrations, (ii) nonstandardized sensors, e.g. supplied by home-grown weather stations, as well as their mobile phones which can be fitted with basic sensor systems (measuring air pressure for example) and finally, (iii) subjective observations from the end users themselves in the form of their personal perceptions on either the environment (weather, pollution, pollen, etc.) or how this affects them directly, such as having a headache, being dizzy, sneezing, etc. Using all of the data generated by this heterogeneous mix of sensors is not straightforward. What exactly was measured? Who measured it? How reliable is the data and how can it be accessed? The required sensor data is often only available through proprietary, non-standard interfaces, complicating the use of data from different data providers in the same model.

To make use of these heterogeneous data specialized models are required. These models are created by domain experts [17]. Since domain experts are not

programmers, and since mathematical modelling is best done with specialized tools, the models needed for processing the available data are often created in packages like MATLAB or R. As a result the models are not easily deployable as a service on the internet and thus not available to (client) applications.

The Open Geospatial Consortium (OGC) is a non-profit, international voluntary consensus standards organization that is leading the development of standards for geospatial and location-based services [7]. The standards are applied to a wide range of domains. The primary objective of the standards is to support information discovery, access, processing, fusion and decision-making. The relevant OGC standards with respect to sensors and sensor data are the OGC Sensor Web Enablement (SWE) standards, which allow sensor data to be accessed in a uniform way over the Internet with web services and XML encodings.

The SWE services of relevance to this paper are: (a) Sensor Observation Service (SOS) to access sensor observations with queries filtering on the phenomenon (property) and the spatial and temporal domains of the observations, (b) Sensor Planning Service (SPS) to parametrise, schedule and execute assets such as sensors. The main operations of the SPS are DescribeTasking (to get the tasking parameters), GetFeasibility (to ascertain if the asset can be tasked with the given parameters) and Submit (to actually execute the task). During the execution the operations GetStatus and Cancel are available.

The SWE services are accompanied by two fundamental information models:

1. The Sensor Model Language (SensorML) which describes the static metadata of a sensor such as make, model, operator and measurement procedure. It can also describe the parameters that a sensor accepts if a sensor can be tasked using a SPS request.
2. The Observation & Measurement Model (O&M Model) which describes observations and the relation between an observation, the associated feature of interest (what was observed) and the associated sensor (how it was observed).

To allow the publication of models as a service, with a well-defined interface, Fraunhofer IOSB has designed the Fusion4Decision architecture. In this architecture, different types of models, called Fusors, implemented in languages like R, MATLAB, Python or C, can be started and tracked using a SPS and the results of the model execution can be retrieved using a SOS. In other words, the model is treated as a sensor and the model results are considered to be new observations. When an application needs a certain model to run on a specific set of data, the application can task the F4D SPS server to execute the model. During the execution of the model the application can track the progress of the execution. Once the model has finished, the application can access the resulting observation through the SOS and use it for further processing, or display it to the user. The requirement for the Fusors is that they are implemented in a way so they can be started by a Java application, that they accept their parameters in a single SensorML document and that they return a single observation document. The returned document can link to other generated files for increased efficiency.
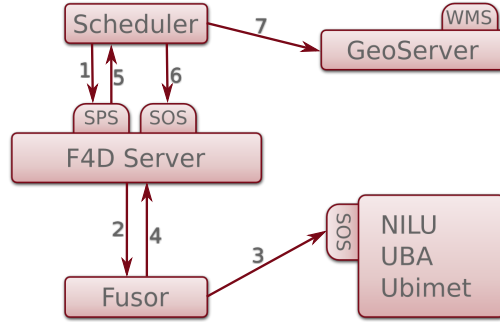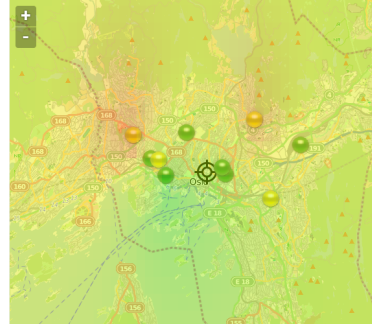
Fig. 5. Use of the F4D architecture in ENVIROFI

In the current prototype of the PEIS the Fusion4Decision system was successfully used to turn point measurements of air pollutants, gathered at fixed measuring stations, into overlay maps of pollutant concentrations covering the cities of Oslo and Vienna. Figure 5 describes this process in detail. When the client that wants to schedule a task (the Scheduler) determines a new model execution is needed, it requests a new task by calling the GetFeasibility and Submit operations of the SPS interface of the F4D server (1). The F4D Server starts the requested Fusor with the supplied parameters (2). The Fusor requests data from various SOS servers (3), and calculates the requested results. The Fusor returns the results to the F4D server (4) which makes them available in its SOS interface. The F4D server then sends a notification to the client that scheduled the task (5) or the client that regularly polls the F4D server to monitor the progress of the execution of the task using the GetStatus operation. The Scheduler then fetches the results from the SOS service of the F4D server (6) and pushes the results into the GeoServer (7), such that they become available for the PEIS to use.

Fig. 6. A map of Oslo with an overlay showing the concentration, interpolated from measurements at the 10 stations shown

The currently used underlying models use Bayesian maximum entropy calculations to interpolate sensor data from the different measurement stations (Fig. 6). The Bayesian Maximum Entropy method is able to consider soft sensor data (e.g. the sensor value lies in an interval) and additional phenomenological knowledge in the form of models. The results are statistics quantifying the uncertainty of the spatiotemporal interpolation given the uncertainty of the available information.

The overall BME fusion method is structured in three stages [8, 9]:

1. prior stage: Consideration of general physical and scientific knowledge $\mathcal{G}$ about the spatio-temporal properties of the phenomenon of interest. It is what is known before experience with the specific situation is applied. The prior probability distribution of the so-called random field of the phenomenon is determined using the maximum entropy (ME) principle, i.e. it is the most uninformative (unbiased) probability distribution given only $\mathcal{G}$.
2. meta-prior stage: Consideration of case-specific hard and soft data of the phenomenon of interest. This information is denoted by $\mathcal{S}$ (for specific knowledge) and is based on observations and measurements. Hard data refers to values believed to be accurate. Soft data is accompanied by uncertainty information such as a probability distribution for the value range.
3. posterior stage: Processing (fusion) of the available knowledge $\mathcal{G}$ and $\mathcal{S}$ of the prior and meta-prior stages respectively to make a probabilistic map of the phenomenon for a given set of spatiotemporal points (typically a grid). The map is a statement of the general knowledge $\mathcal{G}$ relative to the case-specific knowledge $\mathcal{S}$ and is derived using Bayesian conditional probabilities.

The knowledge $\mathcal{S}$ is represented as an observation collection described with the O&M model and including uncertainty information in UncertML. The map resulting from the posterior stage is represented as a coverage with associated uncertainty information. The BME models are implemented in MATLAB and compiled as a Java program for integration into the F4D architecture. Models written in other languages, such as R, can be integrated as well.

5   Conclusions and Recommendations

Future work will include combining these personal health reports with atmospheric condition measurements to create personalized health profiles. This will help to develop a closer relationship between PEIS users and their environment, from both, an atmospheric observation and health observation perspective. Raising public consciousness of their environment and its potential impact upon them is a key goal of the project as this could lead to improved decision making on the side of the users, which will lead to improved health and better knowledge in turn. These objectives are just the tip of the iceberg, however, since for systems like the PEIS that can include sensors at the user level, the potentiality is huge. Crowd sourcing, humans as sensors and even bio-monitoring is still in its infancy and the future will see more and more micro sensors deployed and contributing to our understanding of both the environment and human health impact. The PEIS and the Fusion4Decision system that supports it will play a role in this future as a platform for data collection, processing and dissemination.

Acknowledgments

References

1. Apache couchDB. http://couchdb.apache.org/ (Apr 2013)
2. ENVIROFI: The environmental observation web and its service applications within the future internet. http://www.envirofi.eu/ (jun 2013)
3. PhoneGap. http://phonegap.com/ (Apr 2013)
4. Sencha touch: Build mobile web apps with html5. http://www.sencha.com/products/touch (Apr 2013)
5. AIRNow: Air quality index (aqi) - a guide to air quality and your health. http://www.airnow.gov/index.cfm?action=aqibasics.index (Apr 2013)
6. Anderson, C., Lehnardt, J., Slater, N.: CouchDB: The Definitive Guide: The Definitive Guide. O'Reilly Media (2010)
7. Botts, M., Percivall, G., Reed, C., Davidson, J.: Ogc® sensor web enablement: Overview and high level architecture. In: GeoSensor networks, pp. 175–190. Springer (2008)
8. Christakos, G.: Modern spatiotemporal geostatistics. Oxford University Press, New York, 2nd edn. (2001)
9. Christakos, G., Bogeart, P., Serre, M.L.: Temporal GIS: advanced functions for field-based applications. Springer (2002), ISBN 3540414762, 9783540414766
10. FI-PPP: the future internet public-private partnership programme. http://www.fi-ppp.eu/ (Apr 2013)
11. Havlik, D., Kutschera, P., Geyer, J., Egly, M.: Geospatial service interface and encoding for mobile applications. In: Arndt, H.K., Knetsch, G., Pillmann, W. (eds.) Man-Environment-Bauhaus: Light up the Ideas of Environmental Informatics, Proceedings of 26th International Conference on Informatics for Environmental Protection. pp. 413–421. Dessau, Germany (2012), ISBN: 978-3-8440-1248-4
12. Havlik, D., Schade, S., Sabeur, Z.A., Mazzetti, P., Watson, K., Berre, A.J., Lorenzo Mon, J.: From sensor to observation web with environmental enablers in the future internet. Sensors 11(4), 3874–3907 (2011)
13. Klopfer, M. (ed.): EO2HEAVEN, Mitigating Health Risks. The EO2HEAVEN Consortium http://eo2heaven.org (2013)
14. Kobernus, M., Havlik, D., Van der Schaaf, H., Pielorz, J., Falgenhauer, M.: A quest for affordable personalized atmospheric exposure estimates. Proceedings of the 26th International Conference on Informatics for Environmental Protection (2012)
15. Kobernus, M., Pielorz, J.: ENVIROFI D2.3.2: Functional and Organisational Specification for PIS Pilot II. http://envirofi.eu/Downloads/PublicDeliverables/tabid/4983/Default.aspx (May 2012)
16. Schade, S., Fogarty, B., Kobernus, M., Schleidt, K., Gaughan, P., Mazzetti, P., Berre, A.: Environmental information systems on the internet - a need for change. In: Hrebicek, J., Schimak, G., Denzer, R. (eds.) Environmental Software Systems. Frameworks of eEnvironment, Proceedings of the 9th IFIP WG 5.11 International Symposium, ISESS 2011. pp. 144–153. Brno, Czech Republic (2011)
17. Watson, V., Watson, K.: Design of a software framework based on geospatial standards to facilitate environmental modelling workflows. In: Seppelt, R., Voinov, A.A., Lange, S., Bankamp, D. (eds.) 6th International Congress on Environmental Modelling and Software (iEMSs). iEMSs (July 2012), http://www.iemss.org/society/index.php/iemss-2012-proceedings. ISBN: 978-88-9035-742-8